DOCUMENT RESUME

ED 166 217 TM 007 920

AUTHOR Haag, Carl H.
TITLE The Uses and Abuses of Standardized Testing.
PUB DATE 78
NOTE 7p.; Paper presented at the Summer Convention of the American Association of School Administrators (Minneapolis, Minnesota, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; City Wide Programs; Educational Accountability; *Educational Assessment; Educational Objectives; *Educational Testing; Elementary Secondary Education; *Standardized Tests; State Programs; Testing Programs; Test Interpretation

ABSTRACT

Some uses of achievement tests to provide information for school administrators are described. Three systematic programs are described briefly: (1) a computer-based approach to management and planning designed to answer what is to be learned, how is it to be learned, and has it been learned; (2) a problem solving design for accountability that makes the school the unit of analysis; and (3) a statewide program designed to help classroom teachers, principals, and curriculum directors conduct systematic evaluations. The usefulness of standardized tests for determining how well educational objectives are being met is briefly considered. They are described as probably the most open, most consistent, least biased, cost-effective, and convenient ways of getting information. The importance of a measurement specialist to a school district is mentioned. Comments are made about the importance of using and sharing the information obtained from testing programs. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED166217

THE USES AND ABUSES OF STANDARDIZED TESTING*

All I know about superintendents is what I read in the newspapers. The same holds true for the others who administer schools. Some of you in the audience may feel the same way: that everything you know about tests is what you read in the newspapers. Who would have thought a few years ago that SAT's would make the headlines as they did because of the score decline or that minimum competency testing would be under consideration in every state in the union.

Education, including educational testing, is now very much in the public eye, as you well know. While acknowledging the involvement of legislators wrestling with basic skills, revolting tax payers, and others in educational affairs, it is my intention to focus on aspects of standardized testing which are under the jurisdiction and control of the school district itself.

I am taking this tack because the management of education, like politics, is the art of the possible, and since superintendents and board members have little control over these outside forces, our time can be more effectively spent in the area where you have fundamental responsibility.

I should also admit that it is obvious, from what I read in the newspapers, that you know far more than I about the educational ferment of the age in which we are living. I cannot even point to any scars I bear from my involvement with testing for almost 30 years, particularly during the last half of that period while associated with the Advanced Placement Examinations. These are almost the quintessence of what achievement tests should be. That is, AP exams are designed to measure student achievement in a carefully defined college-level course, they are carefully crafted over a two year period by practicing teachers, they contain both multiple choice and free response or essay questions, they are partly norm-referenced and partly criterion-referenced, and they can be used to provide appropriate recognition for achievement as colleges award exemption and credit.

Further, the AP Program generally seems to have a powerfully positive effect on students, teachers, and schools. Testimonials like the following from a teacher at a typical California High School (average IQ 100, 2750 students, 28% black and Mexican-American) are not uncommon: "I believe (Advanced Placement) to be the most important single program in our department elevating our curriculum standards." That is pretty heady.

To move me into the real world of testiness has required some intense weeks of talking with the many experts among my colleagues at the Educational Testing Service and reading a slew of journal articles. Then, as an administrator myself, I tried to put the information together in a way which might be most usable to you. Three themes will be shared with you:

1. If you don't know where you're going, any road will take you there. This has to do with setting educational objectives.

2. The unexamined life is not worth living. This has to do with obtaining dependable information via tests and other assessment devices.

TM007 920

*Dr. Carl H. Haag, Director, Advanced Placement Program, Educational Testing Service, Princeton, New Jersey AASA National Center for the Improvement of Learning, Summer Convention/Minneapolis/1978

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carl H. Haag

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

3. Do unto others. This has to do with the obligation to take action on the basis of your goals and the information you have at hand.

"If you don't know where you're going, any road will take you there" would seem to be stating the obvious: you need to have goals and objectives before considering any form of educational assessment. It may be obvious but it has not always been observed. Over the last 50 years batteries of standardized tests placed in a number of school systems without sufficient rhyme or reason and of course without much payoff. This makes little sense, and it does cost dollars.

The assumption being made is that educators have goals they wish to achieve, they will take action to achieve their goals, and they need reliable information-including tests-to see how well they are doing.

There are three identifiable groups each requiring different kinds of information and possibly requiring different kinds of tests as well.

Superintendents, School Board Members, other administrators need ways of assessing educational programs. Norm-referenced achievement test batteries and possibly aptitude tests are helpful for this purpose.

Teachers - need ways of evaluating achievement and diagnosing special needs of students. Criterion-referenced achievement tests and diagnostic tests can be used by them.

Counselors - need ways of obtaining information about general abilities, interests, special aptitudes or skills of students, possibly attitudes. Tests and inventories measuring these traits are tools available to counselors.

Again, I would like to narrow our focus to the use of tests to serve the first group, the needs of superintendents and board members in program evaluation. What kind of goals might be appropriate for program evaluation? They might be as clear and direct as determining how district student achievement looks across time. Is achievement in reading, mathematics, and other subjects stable, improving, getting worse? Is it currently at a level which seems desirable for your school district? How does achievement within the district compare with that of other schools in the state or the nation and is it reasonable that your district is lower, at the same level, or higher? District-wide norm-referenced testing provides data which allows you to begin to answer questions like this, and to go on from there to determine what you might do if you are not achieving your desired goals.

At the other extreme there are now some well-laid out, comprehensive, systematic approaches to program planning and evaluation. From my perspective, this type of approach has much to commend it, particularly within the current climate, because it provides a framework and a format for those in a school district to agree on what to do, how they should do it, and whether they have done it. Let me try to convey something about this approach, which seems to be gaining increasing numbers of adherents, by describing three examples: one a school district which developed their own scheme, a second that turned to a model developed by outside consultants, and a third which was a state-wide development.

1. The tri-county Metropolitan Portland, Oregon school districts started to develop a computer-based approach to management and planning in the

early 70's. Their mission was to answer three questions: "What is to be learned? How is it to be learned? Has it actually been learned?" There were also six needs that provided goal targets:

1. Participation of students, parents, teachers, school boards, and local community citizens in decisions about what the schools should teach.

2. Individualization of instruction

3. Cross-disciplinary education

4. Accountability

5. Long-range planning and systematic control of educational development.

6. Effective teacher evaluation

The general approach is described, under Model I, in the National Council on Measurement in Education (NCME) publication "Measurement in Education" which is listed on the sheet available in this room. There may be individuals from Portland in the audience who can tell us what has happened more recently.

2. In New York City a problem-solving design for accountability in education is being developed following a model developed by Dr. Henry Dyer of ETS in the late sixties and early seventies. Before going further, I want you to understand that I am talking about New York City and about accountabilty; the fact that this approach has been endorsed by the Superintendent, the Board of Education, citizen groups, and the teachers' union speaks volumes about how well it has been worked out.

Dyer has set forth three general principles: (1)"the professional staff of a school is to be held collectively responsible for knowing as much as it can (a) about the intellectual and personal-social development of the pupils in its charge and (b) about the conditions and educational services that may be facilitating or impeding the pupil's development." (2)They should be "held collectively responsible for using this knowledge as best they can to maximize the development of their pupils toward certain clearly defined and agreed-upon pupil performance objectives." (3)The board of education has a corresponding responsibility to provide the means and technical assistance whereby staff of each school can acquire, interpret, and use the information necessary for carrying out the two fore-going functions." Finally, Dyer believes "To hold teachers, or anybody else, accountable for some sort of "guaranteed pupil performance" is likely to do more harm than good in the lives of the children." He identified four variables that must be taken into account in any thorough accountability research design:

1. Input variables, or the characteristics of the pupils as they enter a particular phase of schooling-their health, level of achievement, self-concept, aspirations, and other considerations

2. Surrounding environmental conditions within which the school operates including home, community, and school conditions

3. The educational process

4. Output variables or the characteristics of the pupils as they emerge from a particular phase of schooling

The first two variables are difficult to take into account and yet, as Coleman has demonstrated, they may be responsible for a large proportion of the variability in performance levels. I have seen estimates that range from 50% to 95%-the latter leaves the school with no significant function.

The New York City design handles this problem in a brilliantly simple way. They have made the school the basic unit of analysis, and they have done so because all the complex factors within the input variables and the surrounding environmental conditions and even the educational process operate, on, within, and through a particular school. And how do they handle their assessment? Simply by obtaining two measurements for each school-for example, achievement in reading at the beginning of the third grade and at the beginning of the fourth grade. From this they are able to make two kinds of analyses:

1. The percentage of students in the school below minimum standards in a subject such as reading.

2. The "Student Development Index" which shows how much progress each school has made with its children in the past year.

The concern for educational accountability in New York City is focused on what the school can do. How well it works will be known in the next few years. More information about this approach can be found in the Dyer Phi Delta Kappan article and in the McDonald and Forehand New York University Education Quarterly article listed on your sheet under Model II.

3. The third Program evaluation model was developed by the California State Department of Education. In the Foreward to the Program Evaluator's Guide, Wilson Riles states-and remember this was well before Proposition 13-"Evaluation of school programs is becoming more of a necessity for survival than a luxury enjoyed only by affluent districts. As financial resources diminish, decisions on how to allocate the available funds must be made." California set about to develop a training program to help classroom teachers, principals, and curriculum directors learn how to conduct systematic evaluations of a local school or classroom program.

It would be easy to dwell at length on the impressive nature of the Guide and the accompanying Workbook. They are like the Whole Earth Catalog to program evaluation. The section in the Guide on "Selecting or Developing Assessment Instruments" is one of the most succinct on-target presentations I have yet seen. Since these materials are published by ETS, you can leaf through the books at the ETS exhibit or you can take one of the descriptive leaflets available in this room.

Now that I have shared some maps to help you determine where you might go and which road will take you there, we can turn to the second theme: "The unexamined life is not worth living." Obviously, the obverse that Plato was stating is that the examined life is worth living. Now, it can be suggested that, for educators, standarized tests are a means of examining life. Let's briefly consider standardized tests, thinking about them as tools which you can use to provide information as to how your educational objectives are being met.

Before going further it is important to state that tests are not the only instruments which can be used for assessment nor should they be the only source of data for decision making. These other assessment instruments include questionnaires, observational records, and logs maintained by students, teachers, and the school itself in its own records. But it is likely that these devices will be used in addition to tests because of the special properties tests have. One way we could find out the importance of tests would be to suppose that standardized tests were controlled by the Organization of Test Exporting Countries and that OTEC decided to place an embargo on all tests. Of course it would not be as critical as oil, but I suspect the disruption in education would be severe and the protests loud.

Why are standardized tests of such importance? Possibly for these three reasons: they are among the few means we have of obtaining data, they are probably the most open, most consistent, least biased of the available measurement devices, and they are cost-effective, convenient, and simple to employ. The limited ways-questionnaires, observational records, logs-we have of getting useful information is self-evident, and there is not much more that need be said about the convenience or the costs of tests, which represents only one-tenth of 1 percent of the total instructional budget of the average school district. Their openness, consistency, and lower level of bias deserves further comment.

The standards implicit in "standardized" achievement tests are in the construction, in the administration, and in the interpretation. Standardized tests are constructed to meet specifications which educators have agreed are appropriate within the discipline and within the subject itself. The specifications, or educational objectives, are out in the open, as are the questions on the tests themselves. And these properties are subjected to pretests, statistical analysis, validation, and review-all of which help verify that they will measure what is intended, will do it with known accuracy, and will not reflect the unknown, bias that occurs when people make judgments in other ways.

Most people, I suspect, without knowing a great deal about the technical aspects of test development, trust tests. And, generally, they should. In fact, the problem may be that many people trust tests to provide more information than is at all reasonable. The SAT is not the Dow-Jones Index of American Education, a social indicator telling us about academic aptitude across the nation or at a particular school. Nor does a standardized achievement test tell us more about a school than what is defined in the test specifications. And now we are at the heart of it: it is the interpretation of test scores that can cause problems.

Opening up the discussion of test interpretation could carry us through the rest of the day. You will be spared this, because there are three thoughts which can bridge over the complexities involved in interpreting tests and which summarize ideas you can use.

A. You probably will want to have, for norm-referenced achievement tests, normative information-that is, the test results obtained by defined groups of students-on national, state, and local groups.

B. All current methods of reporting scores-percentiles, standard scores, stanines, and grade equivalents-present sophisticated problems

of interpretation and are likely to be misunderstood, even by teachers. There is evidence that teachers do not learn a great deal about tests and measurement within the typical teacher training program.

C. The single most effective action any school district can take, assuming it will be gathering data by using tests and other instruments, is to appoint at least one measurement specialist, with a sufficient budget to work with. A Director of Research and Evaluation can provide the expertise needed to help select appropriate tests and to help interpret them.

Now, we have reached the point where you have information which should tell you how well you have reached the goals and objectives you have set for yourself. And this brings us to the concluding thought, the half Golden Rule: "Do Unto Others."

"Do Unto Others" is meant to suggest two things. The first is that something should be done with the information that has been obtained, something for students, something that will help them in their education. Otherwise, it would be like designing an elegant new sailing ship, checking how it functions, but not using the data from the test runs to make modifications before it actually competes in races. Using the information at hand means making an investment of time, effort, and funds, it may mean obtaining more test information about certain students, and it probably will mean finding ways to determine whether anything happened as a result of the action you take.

The second suggestion embedded in "Do Unto Others" might be better as stated as "Give unto others," that is, to share information we have with others. Perhaps it reflects some sort of societal ambivalence that we are concerned both about individual privacy and full disclosure at this period of our existence. Fortunately, it is possible to make information about persons "privileged" while presenting information about groups of individuals "publicly." Needless to say, the interpretation of tests or other kinds of information to the public is, as the understatement of the day puts it, "an interesting challenge." However, if you have done your homework, or more appropriately your schoolwork, and have, in a reasonably organized fashion, set your objectives, have found out how well you are meeting them, and have taken action to do better, the next time I read about you in the newspaper, the news will be of your making and it will be good!